

Biomedical Data Science Online Curriculum on HarvardX

Harvard University

PI: Rafael Angel Irizarry

Grant Number: 1R25GM114818-01

Unprecedented advances in digital technology during the second half of the 20th century have produced a Big Data revolution that is transforming science, including health and biomedical research. Scientific fields that have traditionally relied upon simple data analysis techniques of smaller datasets have been transformed recently by technologies that continue to expand the possibilities of observing and deciphering molecular entities in an unprecedented way. However, training for the necessary skills and knowledge bases needed to fully leverage big data has lagged behind. The Departments of Biostatistics, Computer Science, and Statistics at Harvard University are partnering with Harvard's Massively Open Online Course (MOOC) initiative, HarvardX, to propose the development of a Biomedical Data Science Online Curriculum. Through this partnership we plan to develop a rigorous and practical curriculum in this nascent field. The overall objective of the proposed research education program is to help prepare the biomedical research community for the Big Data revolution. To accomplish this, we will develop a modular online education program that brings together concepts from Statistics, Computer Science and Software Engineering. Our curriculum will be motivated by real world problems and will serve a wide variety of students with different backgrounds and data analytic needs. Its centerpiece will be a course dedicated to case studies from genomics, imaging and electronic medical records. The case studies will not be artificial in any way and will include all the nuances and grind work associated with modern data analysis. Our specific aims will include: 1) develop and teach an online Biomedical Data Science Curriculum, 2) make the curriculum available in ongoing fashion via the open source edX platform, and 3) disseminate the knowledge gained from preparing and teaching this curriculum. We have put together a team from across Harvard that includes the developers of Harvard's first Data Science class, the faculty of HarvardX's two data analysis online courses, and faculty with expertise analyzing biomedical big data. This team will collaborate to develop a modular, yet fully integrated, set of focused mini- lectures and assessments that will serve as a model for future massively open, self-access online curricula.

PUBLIC HEALTH RELEVANCE: Scientific fields that have traditionally relied upon simple data analysis techniques have been revolutionized by technologies that have massively expanded ways to observe and decipher molecular and physiological entities in an unprecedented way. However, training that is necessary to leverage Big Data has severely lagged behind the technology itself, and bottlenecks in research productivity are evident. To help change this, a Biomedical Data Science Online Curriculum is proposed as a collaborative partnership between the Departments of Biostatistics in the Harvard School of Public Health, Computer Science in the Harvard School of Engineering and Applied Sciences, Statistics in the Harvard Faculty of Arts and Sciences, and Harvard's Massively Open Online Course (MOOC) initiative, HarvardX.